2012

The Effect of Blinded Experts on Jurors' Verdicts

Christopher Robertson

Journal of Empirical Legal Studies
Volume 9, Issue 4, 765–794, December 2012

The Effect of Blinded Experts on Juror Verdicts

*Christopher T. Robertson and David V. Yokum**

"Blind expertise" has been proposed as an institutional solution to the problem of bias in expert witness testimony in litigation (Robertson 2010). At the request of a litigant, an intermediary selects a qualified expert and pays the expert to review a case without knowing which side requested the opinion. This article reports an experiment that tests the hypothesis that, compared to traditional experts, such "blinded experts" will be more persuasive to jurors. A national sample of mock jurors ($N = 275$) watched an online video of a staged medical malpractice trial, including testimony from two medical experts, one of whom (or neither, in the control condition) was randomly assigned to be a blind expert. We also manipulated whether the judge provided a special jury instruction explaining the blinding concept. Descriptively, the data suggest juror reluctance to impose liability. Despite an experimental design that included negligent medical care, only 46 percent of the jurors found negligence in the control condition, which represents the status quo. Blind experts, testifying on either side, were perceived as significantly more credible, and were more highly persuasive, in that they doubled (or halved) the odds of a favorable verdict, and increased (or decreased) simulated damages awards by over $100,000. The increased damages award appears to be due to jurors hedging their damages awards, which interacted with the blind expert as a driver of certainty. Use of a blind expert may be a rational strategy for litigants, even without judicial intervention in the form of special jury instructions or otherwise.

I. Background

The U.S. legal system tasks judges and jurors—both laypersons as to science—with resolving highly technical questions. These laypersons are asked, for example, to evaluate DNA evidence to determine whether it inculpates a particular defendant, to determine the standard of care for lumbar radiculopathy, to interpret epidemiological data to determine whether a given chemical causes an observed disease, and to ascertain the state of the art in a patent suit for computer software. Thus, in both civil and criminal litigation, expert

*Address correspondence to Christopher T. Robertson, James E. Rogers College of Law, the University of Arizona, PO Box 210176, Tucson, AZ 85721; email: chris.robertson@law.arizona.edu. Yokum is J.D./Ph.D. candidate, James E. Rogers College of Law, Department of Psychology, The University of Arizona.

Thanks to Gregory Schwartz, Aaron Kesselheim, and Tom Mauet for serving as actors and consultants; to James Greiner, George Kimball Smith, and anonymous reviewers for the Conference on Empirical Legal Studies and *Journal of Empirical Legal Studies* for commenting on drafts, and to Germar Townsend, Tess Gemberling, Carol Ward, Judy Parker, Barbara Lopez, and Bert Sky for excellent research and administrative support.

witnesses play an increasingly prominent and important role to inform those layperson decisionmakers. In civil litigation, experts appear in the vast majority of trials (Gross 1991), and courts have held in the criminal context that the right to effective counsel includes the right to funding for expert testimony (*Ake v. Oklahoma*).

In the U.S. legal system, expert witnesses are hand-picked by the litigants, coached toward favorable opinions, and compensated for continuing work only as long as those opinions are favorable (Robertson 2010). These experts are thus subject to various cognitive biases, which have been extensively documented in the social sciences (Risinger et al. 2002). Reflecting on his own experience, John Langbein (1985) has explained that "those of us who serve as expert witnesses are known as 'saxophones[,]' . . . a musical instrument on which the lawyer sounds the desired notes." Courts have criticized experts who altogether "cast aside [their] scholar's mantle and [became] a shill" for the party that retained them (*Mid-State Fertilizer Co. v. Exchange National Bank of Chicago*). The problem is not limited to outliers, however. In one study, both federal judges and attorneys cited "experts abandon[ing] objectivity and becom[ing] advocates for the side that hired them" as the most frequent problem with expert testimony (Johnson et al. 2000).

Experimental studies have shown that witnesses interviewed by partisan attorneys prior to testifying tend to deliver more biased testimony than those interviewed by nonpartisan attorneys (Sheppard & Vidmar 1980). In criminal cases, scholars have documented a significant bias of expert opinions in the direction of the side that requested the opinion (Otto 1989). One litigation-oriented study (Gitlin et al. 2004) compared the opinions rendered by radiologists retained by plaintiffs' attorneys in asbestos cases to opinions rendered by a panel of "independent" radiologists retained by defense attorneys.[1] The plaintiffs' experts found physiological abnormalities 95.9 percent of the time, while "independent" reviewers found abnormalities only 4.5 percent of the time. These sorts of biases may operate subconsciously (e.g., an anchoring heuristic driven by the litigant's suggestion as to the amount of damages) or consciously (e.g., a desire to maintain a stream of income from the litigant).

Expert biases may be the subject of cross-examination in trials, but that is typically a zero-sum game, since both litigants tend to exploit these biases. Although court appointment of experts is a possible solution, it is almost never actually employed, for cultural and economic reasons, in the adversarial U.S. legal system (Gross 1991; Cecil & Willging 1993). Without a real alternative to biased expert witnesses, it is not practicable for judges to simply exclude all such testimony; instead, the exclusion doctrine—based on *Daubert v. Merrell Dow Pharmaceuticals*—operates only in marginal cases. It is not a solution for the everyday instances of litigant-induced biases.

When layperson factfinders (whether judge or jury) are left to rely on traditional expert witnesses who are subject to these biases, the factfinders receive relatively weak

---

[1]We put "independent" in scare quotes because the study authors admit that the research was done on behalf of defense attorneys. See Oliver et al. (2004), arguing that the Gitlin "study was done at the behest of attorneys for defendants in asbestos litigation."

epistemic signals as to the scientific truth.[2] Even where there is a realm of legitimate disagreement in a field of expertise, the litigation factfinder only sees hand-picked exemplars from each polar extreme of opinion, and is thus unable to distinguish whether the true distribution of opinion is a 50–50 split or skewed 99–1. This is a selection bias (Robertson 2010:184–85).

Jurors "come to understand the adversary system and on the whole evaluate expert witnesses in the light of this perspective" (Vidmar 1995:173). When faced with two counterpoised and seemingly biased experts, the factfinders sometimes rely on disparities in credibility to prefer one over the other, but they often feel, and sometimes say, that "the two sides have canceled each other out" (Liptak 2008). Notwithstanding the tens of thousands of dollars each side spent on expert witness fees and the hours of trial time consumed by their testimony, it is as if no expert appeared at all (Brekke & Borgida 1988). This epistemic equipoise, resulting from counterpoised hand-picked experts, may be one reason that jury verdicts result in "discordant outcomes" that contradict the views of independent experts who review the merits of the case (Studdert & Mello 2007).

As a consequence, to the extent that factfinders are deprived of meaningful and useful expertise, the quality of litigation outcomes is likely very far from the ideal (Studdert & Mello 2007). The end result is that the deterrence, compensation, and punishment functions of litigation are stymied, and the legitimacy of litigation as a system of dispute resolution is undermined (Solum 2004:190).

More than 60 years ago, biomedical scientists recognized that their objectivity was undermined by extraneous knowledge about which of their subjects were receiving the investigational drug and which were receiving the placebo. Now it goes without saying that "any process using a human as a perceptor, rater, or interpreter should be 'as blind as possible for as long as possible' " (Rosenthal 1978). In recent years, most scientific journal editors have adopted a blind review process to ensure that their editors' decisions are not biased by extraneous information about the authors' identities and institutional affiliations (Snodgrass 2006). On the other hand, of course, readers of biomedical journal articles are not blinded to the authors' identities and may rationally use such information in evaluating scientific research and for allocating their limited time in doing so. For similar reasons, law professors routinely use anonymity in grading (Carrington 1992).

In recent years, several scholars have been developing the idea that blinding could be used by expert witnesses in criminal and civil litigation (see Risinger 2009; Semelka et al. 2010; Robertson 2010; Danaher 2011; Glyptis & Givens 2011). Robertson (2010) provides the most extensive exposition of the concept, explaining that in litigation, blinding could be implemented through the use of an intermediary between litigant and expert. At the request of a single litigant, an intermediary selects a qualified expert and pays the expert to review the case without knowing which side requested the opinion, thus avoiding selection, compensation, and affiliation biases. The fact that an attorney requested such an opinion would not be disclosed to the court or the adversary.

---

[2]According to the standard account, the purpose of trials is so that "the truth may be ascertained." Fed. R. Evid. 102. See Robertson (2010:discussion surrounding notes 16–26).

Under the blind procedure, attorneys would retain ultimate control over their cases, since they individually decide whether to solicit such a blinded expert opinion and then decide whether to call such a blinded expert to trial, after having read his or her expert report. If an opinion turns out to be unfavorable, or not sufficiently strident, an attorney can hide it in "work-product" protection, just as current expert opinions can be hidden. Still, the integrity of the process would be maintained because if an attorney did choose to proceed to trial with a blind expert, he or she would have to disclose how many blinded experts had been consulted on that question, thus allowing the factfinder to evaluate any selection bias. Thus, in practice, litigants would be limited to a single blind expert review on a given question in a case, but would have no risk that an unfavorable blinded opinion could hurt their case. Robertson (2010:209–13) has argued that these secrecy and disclosure rules are already part of attorney work-product doctrine, and a blind expert is fully consistent with the rules of evidence, which allows the blind procedure to proceed without any changes to law, without stipulation of party opposite, and without intervention of the court. It is a litigant-driven solution, designed to leverage rational self-interests.

To our knowledge, blinded experts have to date not been put into practice in litigation. Given that the use of blinded experts seems like it would be a rational strategy for litigants, why have they not already been employed? Attorneys may be unsure of whether jurors sufficiently care about the experts' biases compared to the facts of the case and the story they can tell about those facts. After all, the case is not really about the expert witnesses. Attorneys may be unsure of whether jurors would be able to distinguish between a blinded and an unblinded expert, given that the concept of blinding is somewhat complicated, and the idea of blinded expert witnesses is novel. It may be difficult to present the concept in a way that is both understandable and reassuring to jurors that it is not some sort of sham engineered by the attorneys.

Most fundamentally, attorneys may be unsure whether the jurors would be more likely to render a judgment in favor of the party employing the blind expert. Even if attorneys are hopeful about there being some benefit, they may be unsure of whether the scale of any such benefit would be sufficiently large to offset the cost of buying a blinded opinion that may turn out to be unfavorable and thus unusable. If the benefit provided by a blinded expert is relatively small, an attorney may prefer to proceed with a hand-picked expert, who may also be more skillful as a witness or more malleable in the details of his or her testimony.

Further, attorneys may wonder if it would be necessary to first persuade the judge to provide additional jury instructions explaining and endorsing the concept of blinded experts for it to have any effect. If an attorney needs that additional boost to receive any advantage from the blind procedure, uncertainty about the chances of obtaining that cooperation may also deter attorneys from using the procedure.

Ultimately, these are empirical questions, ones that turn on whether jurors attend to the problem of litigant-induced biases, whether they can understand the blinding process, and whether they find it a compelling means of removing bias. Some extant research sheds light on these questions. Ivkovic and Hans (2003) have provided an excellent review of the literature on juror perceptions of expert bias. For present purposes, we need only highlight the findings that as jurors evaluate expert testimony they are sensitive to the appearance of partiality. Particularly when jurors have trouble understanding the technical substance of a

dispute, they tend to evaluate the credibility of the witnesses as a proxy or heuristic for ascertaining the truth.

In a case study of a complex tort trial, Selvin and Picus (1987:27–28) concluded that, "[c]onfronted with so much complex and confusing information, the jurors tended to evaluate the credibility of these witnesses in large part on their personal characteristics rather than on the information they presented." On the basis of a case study, Sanders (1993) observed that jurors tended to "discount[ ] all expert opinions as testimony of hired guns," but noted that jurors still had views about "relative effectiveness of witnesses." Shuman et al. (1994) performed a survey of jurors, asking them to report what factors they found important in evaluating expert testimony: along with qualifications, familiarity with the case, and quality of reasoning, the jurors said that the appearance of impartiality was an important factor. Vidmar conducted interviews with jurors after they had deliberated on a case involving expert testimony, and concluded that "when there are competent experts on both sides, and they offer contradictory or confusing opinions, jurors may resolve the differences by relying on general impressions of character and veracity" (Vidmar 1995:172).

In a series of three mock jury experiments using a fairly complex trial stimulus, Cooper and Neuhaus manipulated the expert witnesses' pay, credentials, and frequency of testifying. They found (2000:156) that the jurors "thought that the plaintiff's witness was influenced by money as a direct effect of its magnitude," and the effect was clearest for those experts who testified frequently. Mock jurors found (2000:165–66, fig.3) for parties with highly paid experts in only 19 percent of cases and for parties with low-paid experts in 57 percent of cases. In a third condition (2000:166), Cooper and Neuhaus found that the interaction becomes especially strong as the complexity of the trial testimony increases. They suggest that in such challenging situations, jurors may shift from central processing to peripheral processing, wherein they begin relying more on cues as to credibility than the substantive testimony itself. Alternatively, Vidmar and Diamond argue (2001) that if jurors had reasons to doubt the messenger's motives, then the jurors may disregard even that testimony that they understood.

Ivkovic and Hans studied transcripts and structured interviews with 55 jurors on seven cases to explore how they utilized expert testimony. They "conclude[d] that both the characteristics of the expert (the 'messenger') and the substantive and stylistic aspects of the testimony itself (the 'message') contributes significantly to the overall impact of expert testimony on jurors" (2003:443). In particular, "jurors carefully examined and weighed potential motives for bias" (2003:464). Ivkovic and Hans also conducted a written survey of 269 jurors. Of particular interest here, "seven of ten jurors either agreed or strongly agreed with the statement that 'lawyers can always find an expert who will back up their client's point of view, no matter what it is.' Just one of every ten jurors disagreed with this statement" (2003:452). Men were especially likely to agree with this point (82 percent vs. 64 percent, odds of 4.62 vs. 1.78).

In a 2009 study, Brodsky et al. asked mock jurors to evaluate the testimony of two actors playing expert witnesses, presenting themselves as either high or low in "likeability," which they "defined as the degree to which an expert is friendly, respectful, kind, well-mannered, and pleasant." They found that likeability did in fact impact credibility overall, but that jurors distinguished between that factor and both knowledge and confidence

(2009:529). Furthermore, the differences in likeability did not impact juror decisions about whether to sentence the criminal defendant to death versus life without parole.

Brodsky et al. reviewed the research on expert witness credibility and fielded five studies of their own, using mock jurors and videotaped experts to form a "witness credibility scale." Their factor analysis yielded a scale consisting of 20 paired adjectives loaded onto four factors: "'Confidence' was the strongest factor and explained 49.76% of the variance in expert witness credibility. Trustworthiness, likeability, and knowledge added significantly to the analysis and accounted for 9.20, 6.56, and 5.10% of the variance, respectively" (2010:899).

Overall then, it is clear that expert testimony can have a significant impact on trial outcomes, and that jurors are sensitive to various factors, including the appearance of partiality. It remains to be seen, however, whether jurors will respond to the concept of blinding expert witnesses, and whether such a response will be significant enough to motivate trial attorneys to utilize that strategy.

II. Method

A. Stimulus and Design

A two-factor between-subjects design was used, wherein participants watched a video of a medical malpractice trial with two expert witnesses, edited such that neither expert witness was blinded, only the plaintiff's expert was blinded, or only the defendant's expert was blinded. Moreover, for the two conditions with blinded experts, there either were or were not special instructions from the judge explaining the blind expertise concept. Thus, there were five conditions for equally weighted randomization. Participants were subsequently asked to render a verdict, make assessments of the credibility of the expert witnesses, and then complete a variety of demographic questionnaires.

The core of the stimulus, viewed by all participants, was a 35-minute video of a staged medical malpractice trial. The script was written by practicing physicians, who also served as both project consultants and the actors playing the expert witnesses. The scenario concerned the failure of a primary care physician to diagnose a possible case of lumbar radiculopathy and refer the patient to imaging, which allegedly would have allowed timely surgery and avoidance of the permanent disability that the patient now suffers. The primary dispute concerned whether the physician-defendant met the standard of care when, instead of ordering imaging, he simply instructed the patient to take painkillers and return if the pain got worse. The case was designed so that there was a right answer to this question of medical doctrine, one given by a national practice guideline published in the *Annals of Internal Medicine* (Chou et al. 2007). According to that guideline and the stipulated facts, the physician *did* violate the standard of care. To avoid confounding with the variables of interest, this guideline was not introduced in the stimulus trial for the present experiment.[3] It is only a reference point for analysis.

---

[3]National practice guidelines are not, of course, used in many other cases involving witnesses, either because they do not exist, are not determinative, or the parties simply decide not to refer to them (Hyams et al. 1995).

The trial consisted of the following sequence: the trial judge's introduction and preliminary instructions (based on the Revised Arizona Jury Instructions (RAJI)), very brief opening statements from the plaintiff's and the defendant's attorneys, the testimony of plaintiff's expert, the cross-examination of plaintiff's expert, the testimony of defendant's expert, the cross-examination of defendant's expert, very brief closing statements from the plaintiff's and defendant's attorneys, and, lastly, jury instructions from the trial judge (also based on the RAJI). This core video alone constituted the control condition.

The core video was edited to create four additional conditions. First, approximately 10 minutes of extra video footage were distributed across the opening statements, testimony, and cross-examinations, a new redirect of one expert, and closing statements, wherein the attorneys and expert witnesses discussed the concept of blind expertise. Using the same language, either the plaintiff's expert or the defendant's expert was transformed into a blind expert with this additional material, while the actor and substantive testimony remained the same. For example, the opening statement of the plaintiff's [defendant's] attorney, when using the blind expert, contained the following addition:

> The evidence will show that our expert, Dr. Pritchard [Dr. Davidson] is a blind expert, like in a blind taste test. This means two things. First, I did not hand-pick Dr. Pritchard [Dr. Davidson]—he comes from a pool of independent experts. Second, I could not have any influence upon him when he made up his mind about this case, since he did not even know which side was asking. That blind protects his integrity and his objectivity. Compare that to the Defendant's [Plaintiff's] witness, and then you can decide who to trust.

The blind expert explained the blind procedure during the direct examination from the attorney for his side. The witness explained that he was randomly selected by the American Association for the Advancement of Science, and that he reviewed the case without knowing which side asked. The witness also explained the possible sources of bias (selection, affiliation, and compensation) and noted that blinding was routinely used in biomedical science.

During cross-examination, the attorney using the blind expert interrogated the opposing, nonblind expert witness, insinuating that he was biased, as follows:

> Q. So, how many doctors in America do you think could be qualified to review this case?
> A. Oh, I don't know, hundreds, thousands.
> Q. Did you ask Mr. Dobbins why he picked you in particular?
> A. No.
> Q. Do you know why he picked you in particular, rather than hundreds of other doctors?
> A. I assume it was on the basis of a recommendation from one of my colleagues, but I don't know.
> Q. Well, it surely wasn't random, right? Didn't Mr. Dobbins pick you because he was confident that you'd give him a favorable review of this case?
> A. No, I think he just wanted a qualified doctor to review the case, and I reviewed the case and we agreed that it might be reasonable for me to serve.

Note that the nonblind expert firmly denies that he is at all biased, as would be expected in an adversarial trial. Moreover, in the blind expert conditions, the party with the nonblind

expert was given the chance to do a redirect examination, which gave that expert an additional chance to explain that he simply reviewed the facts of the case and rendered an honest opinion.

Q. Doctor, the plaintiff's attorney has tried to suggest that you're biased. Just because I asked you [to] come talk to the jurors today. Is that right?
A. That's right.
Q. But what really informed your decision on this case?
A. I read the report. I read the medical record. I read the report, the complaint from the plaintiff. I talked with the defendant. And, you know, based on my own experience and expertise, I put together my report.
Q. And, I didn't tell you what to put in your report, did I?
A. No, you didn't.
Q. I didn't tell you what to say today, did I?
A. No, you didn't.
Q. Doctor, can you explain to the jury about professionalism and what that means?
A. Sure, I mean, you know, physicians have to adhere to certain standards that are required of members of a professional discipline who have, you know, credentials certified by the state and by the, you know, boards of professional conduct, and, you know, so you have to act with a certain amount of reasonableness and truthfulness.
Q. And have you done that here today?
A. Yes, to the best of my ability.
Q. Now, are there certain codes or rules that you have to follow as a professional?
A. Yes, you mean, there are certain ethical codes and basic standards of conduct.
Q. And are you complying with those in your testimony today?
A. Yes.
Q. Thank you.

In all versions of the video, the judge provided several minutes of general jury instructions, defining medical negligence and other relevant concepts, including a standard charge for the jury to consider the credibility of witnesses. We worried that the attorneys' questions and argument about blinding may not be sufficient to change juror votes significantly. Thus, for two of the experimental conditions, the videos were further edited to add a short segment of special jury instructions from the judge regarding the blind expert to test the hypothesis that this further intervention may make blind expertise more efficacious. In particular, the following language was added:

> I previously instructed that you should consider the potential bias of the witnesses you heard. You have heard testimony from a certain kind of expert witness called a "blind expert," which is a method used to try to minimize bias. In this case, the Plaintiff's [Defendant's] expert, Dr. Pritchard [Dr. Davidson] was blinded. This expert was randomly selected by a neutral third-party from a pool of qualified experts, and [he] rendered an initial opinion about this case without being influenced by either side. Therefore you may find the blind expert more credible and you may decide to give the testimony of that expert additional weight. However, the ultimate decision about the credibility of witnesses is yours.

The end result of all this editing was five video conditions: a control containing the core video stimulus; two videos wherein the plaintiff's expert was a blind expert, either with or without special judge's instructions about blinded experts; and two videos wherein the defendant's expert was a blind expert, either with or without special judge's instructions

about blinded experts. The full videos and transcripts of each condition are available upon request from the authors.

B. Instrument

As noted above, we collected demographic data before showing the trial video to the participants. After the videos, binary verdict judgments were elicited by asking: "Based on the instructions provided by the judge in the video, do you believe that the Plaintiff has proved, by the greater weight of the evidence, that the Defendant committed medical negligence?" Participants responded "yes" or "no."

We also asked respondents: "Based on the evidence you saw, please rate this case on a scale of 1 to 6," where 1 was defined as "clearly not medical negligence" and 6 was defined as "clearly medical negligence." This scalar variable was used for more sensitive data analyses and as a proxy for verdict certainty, in that more certain jurors presumably respond toward the extremes, while equivocal jurors respond toward the middle.

Participants who found medical negligence were further asked how much money the plaintiff should be compensated for pain and suffering. Participants were not asked to decide economic damages because the trial did not present evidence about such damages (for the sake of conserving time, and on the assumption that damages may be stipulated by the parties in such cases). Participants were also asked to type a sentence or two explaining their decisions.

1. Credibility Assessments

To assess the relative credibility of each witness, subjects were asked whether they agreed that the expert witness was *knowledgeable, logical, clear, honest, trustworthy,* and *fair.* For each adjective, there was a four-point-labeled Likert rating (1 = *strongly disagree,* 2 = *disagree,* 3 = *agree,* 4 = *strongly agree*). For analyses, these factors were grouped into two composites; how and why this was done is described more fully below.

C. Participants and Randomization

Jury-eligible adults were recruited for an online mock jury experiment via Amazon Mechanical Turk, for $3 compensation, beginning September 14, 2010. (This platform is increasingly used for social science research and has been validated by comparison to known results; see Paolacci et al. 2010.) Participants were told that the study concerned a mock trial, but were not informed about the hypotheses being tested and were blinded as to their assignment into experimental conditions.

Three-hundred-sixty-four persons consented to participate within the first 48 hours. Fifty-eight persons exited the study before completion, and another 31 persons were excluded for failure to follow instructions (e.g., they completed the study in under 35 minutes, indicating that the 35+ minute trial video was not watched in its entirety). As shown in Table 1, the remaining 275 participants, who constitute the overall sample for analyses purposes, included 195 females and 80 males, aged between 18–77 years ($M = 32.5$

Table 1: Demographics of Experimental Subjects

		BE for Plaintiff (n = *114*)			*BE for Defendant* (n = *100*)				
	Neither BE (n = *61*)	*No Instructions* (n = *54*)	*Instructions* (n = *60*)	*Subtotal*	*No Instructions* (n = *52*)	*Instructions* (n = *48*)	*Subtotal*	*Subject Totals* (N = *275*)	*U.S. Census*
Education									
<HS diploma/GED	2%	2%	3%	3%	4%	0%	2%	2%	18%
HS diploma/GED	16%	17%	22%	19%	17%	17%	17%	18%	30%
Some college/assoc.	43%	43%	32%	37%	17%	33%	35%	38%	27%
College grad.	27%	28%	37%	32%	38%	33%	35%	32%	17%
Graduate degree	13%	11%	7%	9%	6%	17%	11%	11%	10%
Gender									
Male	30%	17%	33%	25.4%	33%	33%	33%	29%	49%
Female	71%	83%	67%	74.6%	67%	67%	67%	71%	51%
Age Groups									
18–24	28%	22%	20%	21.1%	31%	21%	26%	24%	13%
25–34	51%	37%	47%	42.1%	35%	38%	36%	41%	18%
35–44	12%	37%	18%	17.5%	15%	21%	18%	16%	19%
45–59	10%	24%	15%	19.3%	15%	19%	17%	16%	27%
60+	0%	0%	0%	0%	4%	2%	3%	1%	23%
Race									
White	73%	78%	73%	75%	85%	83%	84%	78%	74%
Nonwhite	26%	22%	27%	25%	15%	17%	16%	22%	26%

NOTE: The sample was somewhat more highly educated, more often female, and younger than the U.S. population, but had a similar racial distribution (white/nonwhite) and the differences were successfully randomized across the five experimental conditions. "BE" is an abbreviation for blind expert, and "instructions" refers to a special jury charge on the blind expert concept.

years, $SD = 10.9$ years). The most common ethnicities were Caucasian (78.2 percent), African American (8.7 percent), and Asian (4.4 percent). As for formal education, 2.2 percent reported they had not completed high school, 17.8 percent had a high school diploma, 37.5 percent reported some college, 28.4 percent were college graduates, and 14.1 percent had postgraduate work. All persons consented to participate according to institutional review board standards.

Each participant was then randomly assigned to one of the five experimental video conditions. As shown in Table 1, randomization was successful, with no statistically significant differences on demographic variables across experimental conditions.

III. Results

A. Overall Regression on Verdict

We used binary logistic regression to assess the impact of the experimental manipulations on the likelihood that a participant would render a verdict of medical negligence, while also controlling for demographic variables of gender, age, race, and education. The model contained the following independent variables: PLAINTIFF BE (whether or not the plaintiff had a blind expert), DEFENDANT BE (whether or not the defendant had a blind expert), INSTRUCTIONS (whether or not the judge provided special jury instructions focusing on blinding, in addition to the standard instructions),[4] subject's gender, age, race (nonwhite or white), and education (less than college degree or at least college degree). A hierarchical method of entry was used. The four demographic variables were force entered in the first step, followed by forced entry of PLAINTIFF BE and DEFENDANT BE in the second step, and then forced entry of INSTRUCTIONS in the final step.[5]

The full model containing all predictors was statistically significant, χ^2 (8) = 26.83, $p = 0.001$, indicating that the model was able to successfully distinguish between participants who judged for or against the plaintiff.[6] The model correctly classified 63.3 percent

---

[4]The use of instructions, if it increases the credibility of the blind expert as predicted, would increase or decrease the likelihood of a negligence verdict, depending on whether the plaintiff or defendant, respectively, were the side with the blind expert. As such, the main effect of INSTRUCTIONS would potentially cancel itself out, thereby revealing a misleading null result. To avoid this possibility, the model included the interaction terms of INSTRUCTIONS * PLAINTIFF BE and INSTRUCTIONS * DEFENDANT BE. A different approach is to include the INSTRUCTIONS term but restrict the logistic regression sample to compare the condition with no blind expert to either PLAINTIFF BE or DEFENDANT BE, but not both simultaneously—thus meaning that the INSTRUCTIONS effect relates to PLAINTIFF BE or DEFENDANT BE, respectively. The results are consistent with those reported in Table 2 with the interaction term.

[5]Thus, using this method of entry, the demographic variables are used first, to explain as much variation in outcomes as possible; after that, the experimental variables are used to account for any remaining variance.

[6]Regression results are materially the same when using robust standard errors and classical standard errors (the latter are reported herein). See King and Roberts (2012) (recommending models be tested to ensure that both types of error terms generate the same results, rather than simply reporting robust standard errors).

Table 2: Logistic Regression Predicting the Odds of a Plaintiff Verdict

					95% CI for Odds Ratio		
Predictor	*B*	*S.E.*	*Wald's* χ^2	p	*Lower*	*Odds Ratio*	*Upper*
Age	–0.003	0.012	0.047	0.828	0.97	1.00	1.02
Gender	0.305	0.286	1.142	0.285	0.78	1.36	2.38
Race	–0.423	0.313	1.831	0.176	0.36	0.66	1.21
Education	–0.309	0.260	1.406	0.236	0.44	0.73	1.21
Plaintiff BE	0.778	0.393	3.921	0.048*	1.01	2.18	4.70
Defendant BE	–0.776	0.410	3.579	0.059*	0.21	0.46	1.03
Instructions * Plaintiff BE	–0.167	0.395	0.180	0.672	0.39	0.85	1.83
Instructions * Defendant BE	0.430	0.438	0.960	0.327	0.65	1.54	3.63
Constant	0.130	0.498	0.068	0.794	—	—	—

*$p < 0.05$ or marginally significant.

NOTE: The overall model was statistically significant and the presence of a blind expert more than doubled the odds of winning, for either side. Neither the presence of special jury instructions nor any covariates made a unique, statistically significant contribution to the model. $R^2 = 0.09$ (Cox & Snell), 0.12 (Nagelkerke). Model χ^2 (8) = 26.83, $p = 0.001$.

of cases and, with effect size estimates in the range of 0.09 (Cox and Snell R^2) to 0.12 (Nagelkerke R^2), constitutes a meaningful result.

Only PLAINTIFF BE and DEFENDANT BE made a unique, statistically significant contribution to the model. (See Table 2.) In particular, the odds of a verdict in favor of the plaintiff more than doubled when the plaintiff's expert was a blind expert (odds ratio = 2.18, $p = 0.048$). Likewise, the odds of a verdict in favor of the defendant more than doubled when the defendant's expert was a blind expert (odds ratio = 0.46, $p = 0.059$). These findings affirm our primary hypothesis that the presence of a blind expert would have a large effect on the outcome of trials.

On the other hand, the regression model also shows that the use of special jury instructions did not have a significant effect on verdict, and this was true whether the plaintiff ($p = 0.67$) or defendant ($p = 0.33$) retained the blind expert. This finding disconfirms our secondary hypothesis that this additional intervention by the judge would magnify the effect of the blind expert. Instead, it appears that a purely litigant-driven blinding intervention will suffice.

Because randomization succeeded in distributing covariates across conditions, and these overall regression results show that covariates are not driving our main results, we proceeded to conduct planned comparisons of individual experimental conditions, reported in the following sections. Because comparisons of central tendencies (means and medians) may be easier to interpret and exhibit than regression models, we present such comparisons of primary dependent variables, including verdict and verdict certainty, damages, and expert credibility, using parametric and nonparametric tests of significance.

B. Verdict and Certainty

When neither side had a blind expert—as in a usual trial—most participants rendered a verdict in favor of the physician-defendant, even though the scenario had been designed to

Table 3: Frequency (and Percentage) of Verdicts in Favor of the Defendant and Plaintiff, Along with Mean Certainty Score on Six-Point Scale (and Standard Deviation) by Blind Expert (BE) and Instruction Conditions

		Verdict	
Condition	n	*For Plaintiff (%)*	*Mean Certainty (SD)*
Neither BE (control)	61	28 (45.9)	3.46 (1.51)
BE for plaintiff			
No instructions	54	35 (64.8)	4.11 (1.44)
Instructions	60	36 (60.0)	3.90 (1.57)
Total	114	71 (62.3)	4.00 (1.51)
BE for defendant			
No instructions	52	14 (26.9)	2.90 (1.49)
Instructions	48	17 (35.4)	3.04 (1.71)
Total	100	31 (31.0)	2.97 (1.60)

NOTE: Compared to the control condition where 46 percent of respondents voted for the plaintiff, the presence of a blind expert for the plaintiff increased that rate to 62 percent (a difference of 16 points, $p = 0.042$) and a blind expert for the defendant decreased that rate to 31 percent (a difference of 15 points, $p = 0.037$). Special jury instructions from the judge were associated with statistically insignificant decreases in win rates).

be a case of real medical malpractice that the plaintiff should have won. In this control condition, 46 percent of the jurors found medical negligence. (See Table 3.) Introduction of a blind expert significantly altered this status quo. When a blind expert appeared for the plaintiff (and without special jury instructions), 65 percent of the jurors found negligence, an increase of 19 percent compared to the NEITHER BE control condition. A Pearson's chi-square test reveals this difference to be significant (χ^2 (1) = 4.14, $p = 0.042$). When a blind expert appeared for the defendant (and without special jury instructions), only 27 percent of the jurors found negligence, a decrease of 19 percent compared to the NEITHER BE control condition. This difference is also significant (χ^2 (1) = 3.62, $p = 0.037$). The planned comparison thus further confirmed our primary hypothesis that blind experts would drive outcomes favorable to their sponsors.

As in the logistic regression, our secondary hypothesis about jury instructions was not confirmed in this planned comparison across conditions. Within the blind expert conditions, the use of jury instructions about the blind expert was not associated with increased win rates for the party bringing the blind expert. In fact, the jury instructions were associated with statistically insignificant decreases in win rates (a 5 percent difference when the plaintiff had the blind expert (χ^2 (1) = 0.28, $p = 0.596$), and an 8 percent difference when the defendant had the blind expert (χ^2 (1) = 0.84, $p = 0.359$)).

The six-point scale for how certainly the jurors found medical negligence reveals that in the presence of a blind expert, jurors had stronger opinions about how "clearly" the defendant committed medical negligence. The mean score in the control condition was 3.46. (See Table 3.) This score was higher at 4.00 (indicating more clearly malpractice)

Table 4: Mean and Median Pain and Suffering Awards Including Zeros for Defense Verdict (U.S. Dollars), by Blind Expert (BE) and Instruction, with Standard Deviation and 95 Percent Confidence Interval

Condition	n	*Mean*	*Median*	SD	*CI*
Neither BE	61	278,313	10,000	780,584	199,917
BE for plaintiff					
No instructions	54	370,018	100,000	613,708	167,510
Instructions	60	505,695	72,500	1,338,990	345,897
Total	114	441,427	100,000	1,057,069	196,144
BE for defendant					
No instructions	52	86,599	0	216,363	60,235
Instructions	48	144,583	0	481,914	139,933
Total	100	114,431	0	367,722	72,964

NOTE: The mean award was higher ($p = 0.02$) or lower ($p = 0.01$) than in the control condition by about $163,000, depending on whether the plaintiff or defendant, respectively, retained the blind expert.

when the plaintiff retained a blind expert, and was lower at 2.97 (indicating more clearly *not* malpractice) when the defendant retained a blind expert. These results were statistically significant, given an ordinal regression, χ^2 (4) = 23.6, $p < 0.001$.[7]

C. Monetary Damages and Hedging Thereof

Participants who found medical negligence also determined how much money the plaintiff should be awarded in pain and suffering damages. Recall that participants were not asked to decide economic damages (e.g., medical bills and lost wages). Also, the attorneys did not suggest specific values for pain and suffering awards. Descriptive statistics for damages are listed in Table 4. The data were severely right-skewed, with several far outliers.[8] These outliers might be explained by the fact that jurors were not given any guidance or limitations as to the amount of pain and suffering damages they could award.[9]

---

[7]An ordinal regression was performed rather than a *t* test because the latter presumes interval scores, which are not present in a Likert scale.

[8]In the control condition, the z score for skewness equaled 51.53, which is highly significant ($p < 0.0001$); the skewness z scores were also highly significant in the PLAINTIFF BE ($z_{skew} = 30.63$, $p < 0.0001$) and DEFENDANT BE ($z_{skew} = 23.66$, $p < 0.0001$) conditions.

[9]In the real world of litigation, there is heterogeneity as to whether states allow attorneys to request specific amounts of damages for pain and suffering, and furthermore there is heterogeneity as to whether attorneys actually do so when allowed. See Kahneman et al. (1998) for a general discussion of the difficulty jurors face in mapping judgments of pain and suffering into a numeric dollar amount. See Diamond et al. (2011) for a discussion of whether attorneys are allowed to give reference points for pain and suffering damages, how often attorneys do so when allowed, and how juries react thereto when they are offered. See especially, pp. 6–7, explaining that a few states, such as New Jersey and Pennsylvania, forbid attorneys from asking for a specific amount for pain and suffering because such requests would be of "an arbitrary amount" and thus "highly improper," and p. 20, showing that in a sample of Arizona cases, where

The presence of a blind expert impacted pain and suffering damages, but the presence of special jury instructions on blinding did not further increase that effect. Compared to the $278,313 awarded in the control condition on average, the blind expert for the plaintiff yielded an additional $163,000, totaling $441,427. Symmetrically, when the defendant brought the blind expert, the jury awarded $163,000 less on average, totaling $114,431. These effects are notable, especially since the expert witnesses never directly discuss pain and suffering damages in any of the experimental conditions. Because the data were so significantly skewed, a nonparametric Mann-Whitney test was used. Plaintiffs received significantly more pain and suffering damages when they retained a blind expert ($Mdn = 100{,}000$) relative to the control ($Mdn = 10{,}000$), $U = 2{,}755$, $z = -2.31$, $p = 0.02$, $r = -0.17$. Likewise, defendants paid significantly less pain and suffering expenses ($Mdn = 0$) when they retained a blind expert relative to the control ($Mdn = 10{,}000$), $U = 2{,}379$, $z = -2.55$, $p = 0.01$, $r = -0.20$.

These strong effects, of course, are being driven in large part by the underlying win-rates, in particular the zero damages award that comes with a finding that the defendant is not negligent. Still, the finding highlights the dramatic effect that using a blind expert can have on the economic bottom line for a litigant.

A separate question, however, is whether the blinded expert impacted damages awards, aside from the differential win rates. We did not hypothesize such an effect, since the expert testimony in this case shed light on the question of whether the physician met the standard of care (and thus the likelihood of defense verdict), not the severity of the plaintiff's injury. Nonetheless, to investigate the possibility that the type of expert impacted pain and suffering awards, Table 5 shows the damages awarded in each condition, using data only from those who rendered a verdict in favor of the plaintiff. Even with the defense verdicts excluded, the mean changes appear substantial: plaintiffs with a blind expert received about $90,000 more, and defendants with a blind expert essentially "lost better"—paying about $235,000 less. However, Mann-Whitney tests revealed that the differences were not statistically significant at the 0.05 level.[10] Especially given the high variance in these data, we were not sufficiently powered to test the hypothesis that the presence of a blind expert will impact the amount of noneconomic damages awarded.[11] Future experimenters could, however, explore this dynamic with a larger sample size.

---

attorneys are allowed to name a dollar figure for pain and suffering, they did so in 21 out of 31 cases (68 percent), and defense attorneys conceded some amount of economic damages (contingent on a finding of liability) in 19 out of 33 cases (58 percent). Thus, our trial stimulus—which stipulated economic damages, and thus did not ask the jury for a finding on that item, but asked for a pain and suffering award, without providing a reference point, is not an altogether uncommon scenario. Still, for future experiments, it may be prudent to provide such a reference point, as a way to reduce variance in the data.

[10]NEITHER BE against PLAINTIFF BE: $U = 873$, $z = -0.94$, $p = 0.346$. NEITHER BE against DEFENDANT BE: $U = 365$, $z = -1.05$, $p = 0.293$. Data in all three conditions were again highly skewed: Control $z_{skew} = 7.17$, $p < 0.0001$; PLAINTIFF BE $z_{skew} = 20.32$, $p < 0.0001$; DEFENDANT BE $z_{skew} = 7.67$, $p < 0.0001$.

[11]We confirmed the power calculations provided by an anonymous reviewer, showing that the estimated power is 0.06 and 0.10 for the two conditions comparisons with the control group, which means that we had less than 10 percent chance of obtaining a statistically significant effect of the observed magnitude, given the sample sizes.

Table 5: Pain and Suffering Awards of Those Who Found Negligence (U.S. Dollars)

Condition	n	*Mean*	*Median*	SD
Neither BE	28	588,214	225,000	1,782,276
BE for plaintiff				
No instructions	35	551,600	300,000	696,269
Instructions	36	800,847	500,000	1,667,209
Total	71	677,978	300,000	1,281,020
BE for defendant				
No instructions	14	286,946	125,000	346,574
Instructions	17	408,235	100,000	753,558
Total	31	353,459	100,000	598,888

NOTE: Damages awards varied widely, with apparent changes in mean and median in expected directions (higher when the plaintiff had a blind expert and lower when the defendant did), but not at statistically significant levels.

The variability in the awards was also seemingly affected; in particular, the presence of a blind expert reduced the amount of variability, with the standard deviation being about a half million dollars less than when the plaintiff's expert was blinded, and just over a million dollars less when the defendant's expert was blinded. These observations were, however, short of statistical significance.[12] Nonetheless, such apparent changes in variance are noteworthy. Scholars have observed a similar pattern of "horizontal inequity" across real cases in which plaintiffs with similar injuries received vastly different awards, and have suggested reforms to reduce such variability (e.g., Saks et al. 1997; Kahneman et al. 1998). With a very high standard deviation observed in the control condition (nearly triple the mean), the present experiment replicates this problem in a setting where the plaintiff's injury and circumstances are identical. If confirmed in a larger study, this finding would be contrary to the suggestions that unobserved variation in the plaintiff's circumstances may explain jury variance (Hans & Eisenberg 2011:380).

A more realistic assessment of the impact of such juror decisions on the economic value of cases requires several adjustments to the foregoing data. First, it is likely that during the jury deliberation process, and then again in review of the verdict by the trial and appeals judges using the remittur power and statutory caps on noneconomic damages, extreme individual assessments of liability would yield to more modest judgments.[13] To approximate

[12]A statistical test of variability can be computed by using deviation scores, wherein one calculates the absolute distance between each data point and its mean, and then runs a t test on those deviation scores. See Saks et al. (1997:250). Using this method to test NEITHER BE against PLAINTIFF BE and then against DEFENDANT BE reveals, respectively, $t(97) = 0.05$, $p = 0.96$ and $t(57) = 1.51$, $p = 0.14$.

[13]See Davis (1996) and Diamond and Casper (1992), showing that the best predictor of a jury damages award is the median of individual awards, which necessarily reduces the effect of outliers, and Snyder (2000:320) explaining that use of remittitur has become "pervasive in American law . . . as an all-purpose effort to . . . reduce exorbitant damage awards," and American Medical Association Advocacy Resource Center (2011:1), showing that "close to 30 states have laws in place that limit damages in medical liability actions."

this moderating dynamic, extreme outliers were transformed to be within two standard deviations from the mean. (See Saks et al. (1997:249), justifying the use of a two standard deviations transformation of damages awards over logarithmic transformation or truncation.) Second, we assumed that, on average, a medical malpractice case tried to a jury would be worth about $500,000 in economic damages (primarily for medical expenses and lost earnings), in addition to whatever was awarded in pain and suffering damages. This is a conservative assumption.[14] Finally, we included defense verdicts as zero damages because these would be of primary interest to litigants on both sides. Altogether then, the resulting figure represents the economic value of the case—the overall loss suffered by the defendant going to trial under these conditions or the overall benefit enjoyed by the plaintiff. Aside from litigation costs (not calculated here), this number would represent a potential settlement value for the case.

Figure 1 depicts these results, collapsing across the instructions and no-instructions conditions. The shift in over $100,000 in case value for whichever side brings a blind expert shows that blind experts may be a rational strategy for litigants, and thus a practical reform to the system of civil litigation. When the plaintiff had a blind expert, total awards were significantly greater, relative to the control condition, by about $145,000 on the mean, and $572,000 on the median, $U = 2{,}773$, $z = -2.30$, $p = 0.02$. Likewise, awards were significantly less than in the control condition, by about $106,000 on the mean (with identical medians of $0), when the defendant had a blind expert, $U = 2{,}533$, $z = -2.09$, $p = 0.037$. These results are driven by both the significant differences in win rates across conditions (see Table 4), as well as the large but statistically insignificant differences in the amounts awarded across conditions (see Table 5).

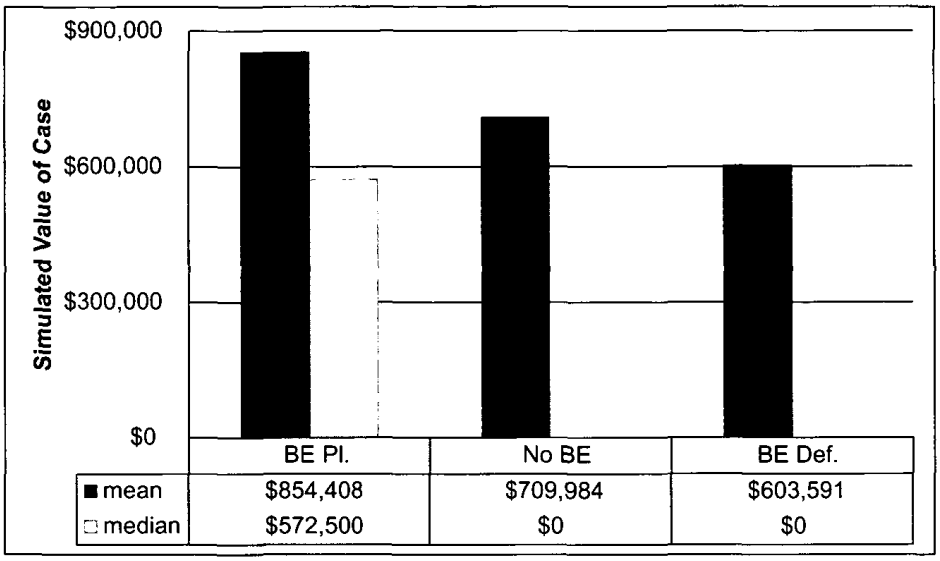
The difference in mean values will be important to rational litigators who are deciding whether to invest in a blind expert opinion ex ante, but the differences in medians may also be particularly attractive given the aversion to losing. In the control condition, representing status quo litigation, most cases are economic losers that are only subsidized by the rare case that wins big. With a blind expert, the median case has positive economic value.

Figure 2 shows mean and median pain and suffering awards, at each level of the jurors' assessments of how "clearly" medical negligence was proven, in all experimental conditions. Note that only those finding negligence are included (thereby censoring levels 1, 2, and 3, where no damages would be awarded, regardless of clearness), and that

---

[14]Given the fixed costs associated with prosecuting a case, plaintiff's attorneys will often decline cases that have modest potential damages. A review of 257 California medical malpractice cases showed mean economic damages awarded by juries of $950,000 and mean noneconomic damages awards of $687,000, for a total of $1,637,000 on the mean (Pace et al. 2004:20, tab. 3.1). After imposition of statutory damages caps, the final mean judgments were $950,000 for economic damages (nearly double what we assumed here), $200,000 for noneconomic damages (about the same as what we found here after adjusting the outliers), equaling $1,150,000 (Pace et al. 2004:23, tab. 3.3). Comparing $1,150,000 to the $709,984 mean simulated award in the control condition here, as shown in Figure 1, shows that altogether our assumptions were conservative. If we had assumed a larger amount of economic damages, or been less aggressive in policing outliers in the noneconomic damages, we could have demonstrated an even larger economic value for the blind procedure.

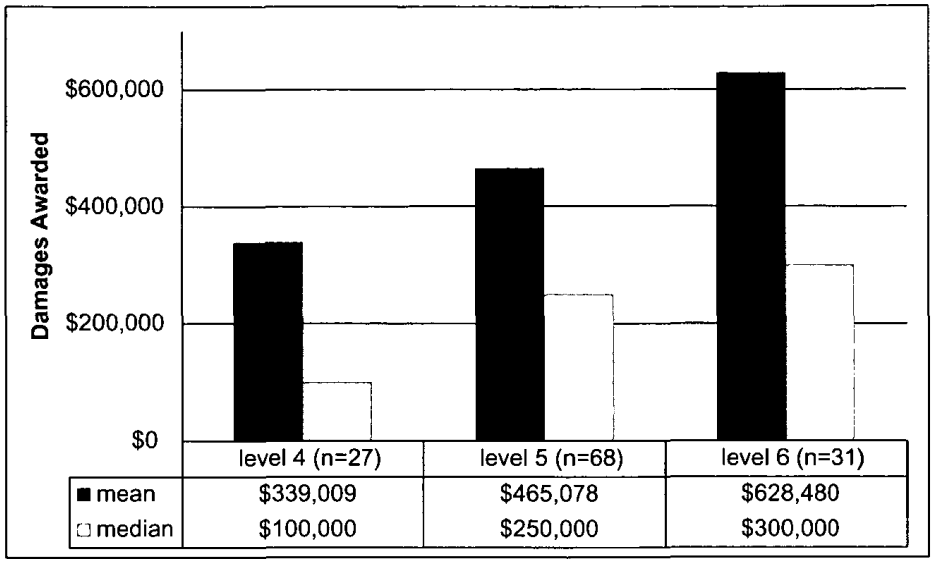
Figure 1: Simulation of economic value of case (U.S. dollars) when neither side (no BE), only the plaintiff (BE Pl.), or only the defendant (BE Def.) has a blind expert, including defense verdicts as zeros. Outlier award values were transformed to within two standard deviations, and $500,000 economic damages were assumed. On these assumptions, the tactic of using a blind expert pays over $100,000 on average to the litigant that uses the tactic, conditional on the expert rendering a favorable, usable opinion not rebutted by a blind expert on the other side.



outliers were transformed to within two standard deviations of the group mean. As can be seen, as verdict certainty increased, the mean and median monetary award also increased. A Kruskal-Wallis test (used because the data were highly skewed) revealed these differences to be significant, χ^2 (4, $N = 130$) = 10.694, $p = 0.030$.

In principle, in cases where the malpractice was in fact more egregious, a plaintiff could experience more pain and suffering, which would thus appropriately support a more generous damages award. Our experiment did not, however, manipulate any substantive testimony about how egregious the malpractice may have been; we simply manipulated the credibility of the witnesses through use of blinded experts. Thus, we think that a more plausible explanation is that jurors responded on the verdict certainty scale in terms of their own epistemic uncertainty about whether malpractice had occurred. They then apparently "hedged" any uncertainty in their verdict by awarding a smaller pain and suffering award than if they were confident it was clearly medical malpractice, where they presumably awarded what they perceived as full compensation. Given that the presence of a blind expert also affected juror certainty levels, this hedging dynamic may explain why the presence of blind expert was correlated with damages awards (see Figure 1).

Figure 2: Mean and median pain and suffering award (U.S. dollars), by levels of medical negligence certainty. Jurors apparently "hedged" by awarding less pain and suffering damages if they found the proof of malpractice less "clear," as reported on a six-point Likert scale. This figure includes all data collapsed across conditions, with outliers transformed to within two standard deviations of the mean.



D. Expert Credibility Assessments

We hypothesized that use of a blind expert would impact verdict outcomes because jurors would perceive those experts as more credible. In this section, we examine overall credibility across conditions, and create two subsidiary constructs to investigate the dynamic more precisely.

Research on perceptions of credibility suggests that credibility is usefully thought of as a construct composed of multiple factors. For example, Brodsky et al. (2010) propose a Witness Credibility Scale composed of the factors *knowledge, likeability, trustworthiness,* and *confidence.* In this study, participants judged each witness in terms of how *knowledgeable, logical, clear, honest, trustworthy,* and *fair* he seemed, using a four-point Likert scale from 1 (strongly disagree) to 4 (strongly agree) for each term. We added the scores on each to create an overall credibility score.

As shown in Table 5, assessments of overall credibility followed the expected pattern. Under the usual circumstances when neither expert was a blind expert (i.e., the NEITHER BE condition), even though the defendant tended to win, the plaintiff's expert was perceived as more credible ($M = 6.39$, $SD = 1.21$) than the defendant's expert ($M = 5.96$, $SD = 1.28$), $t(61) = -2.03$, $p = 0.05$, $r = 0.25$ (see Table 6). This discrepancy in credibility was significantly greater whenever the plaintiff's expert was a blind expert ($M_{\text{plaintiff}} = 6.94$,

Table 6: Mean (Standard Deviation) Assessments of Overall Expert Credibility

	Plaintiff's Expert			*Defendant's Expert*		
	Credibility	*Skill*	*Genuineness*	*Credibility*	*Skill*	*Genuineness*
Neither BE	6.39 (1.21)	3.23 (.062)	3.16 (0.65)	5.96 (1.28)	3.08 (0.61)	2.87 (0.75)
BE for plaintiff	6.94* (1.17)	3.51* (0.58)	3.43* (0.67)	5.35* (1.20)	2.90 (0.63)	2.44* (0.70)
BE for defendant	5.73* (1.14)	3.10 (0.53)	2.63* (0.72)	6.28* (1.07)	3.14* (0.55)	3.14* (0.59)

*Planned contrast relative to NEITHER BE condition is significant at $p < 0.05$.

NOTE: Based on four-point Likert scales from 1 (strongly disagree) to 4 (strongly agree) for each term, subdivided by skill and genuineness, when neither (NEITHER BE), only the plaintiff (BE FOR PLAINTIFF), or only the defendant (BE FOR DEFENDANT) retains a blind expert. Transforming one expert into a blind expert not only increased his own credibility, but also *decreased* the *other* expert's credibility.

$SD = 1.17$, vs. $M_{\text{defendant}} = 5.35$, $SD = 1.20$), $t(114) = -9.48$, $p < 0.001$, $r = 0.66$). On the other hand, when the defendant's expert was a blind expert, the defendant's expert was perceived as significantly more credible ($M = 6.26$, $SD = 1.07$) than the plaintiff's expert ($M = 5.73$, $SD = 1.14$), $t(89) = 3.43$, $p = 0.001$, $r = 0.34$).

Interestingly, transforming one expert into a blind expert not only increased his own credibility, but it also *decreased* the *other* expert's credibility. For example, the credibility of the defendant's expert dropped significantly from a 5.96 (in the NEITHER BE condition) to a 5.35 (in the PLAINTIFF BE condition), when the *plaintiff* had a blind expert).[15] Likewise, the plaintiff's expert's credibility dropped from a 6.39 (NEITHER BE) to a 5.73 (DEFENDANT BE).[16]

Since blinding an expert both enhanced his credibility *and* harmed the credibility of the opposing expert, the end result was an even larger gap in the credibility between the two witnesses. There was only a 0.43 difference in the credibility of the two witnesses in the NEITHER BE condition, but this increased to 1.59 and 0.55 credibility gaps, respectively, in the PLAINTIFF BE and DEFENDANT BE conditions.

To investigate the dynamic more closely, the overall credibility score can be disaggregated into its "skill" and "genuineness" components, which parallel the *knowledge* and *trustworthiness* factors posited by Brodsky et al. (2010). The "skill" construct, consisting of *knowledgeable, logical,* and *clear,* was meant to capture the expertise of the witness in terms of,

[15]Whether the defendant's expert witness was a blind expert had a significant effect on his credibility as perceived by participants, $F(2, 272) = 17.33$, $p < 0.001$. Planned contrasts revealed a significant increase in overall credibility in the DEFENDANT BE condition ($M = 6.28$, $SD = 1.07$) relative to the NEITHER BE condition ($M = 5.96$, $SD = 1.28$), $t(272) = -1.69$, $p = 0.045$ (one-tailed), $r = 0.10$. When only the plaintiff used a blind expert, the defendant's expert witness suffered a worse credibility assessment ($M = 5.35$, $SD = 1.20$) than in the NEITHER BE condition, $t(272) = -3.27$, $p < 0.001$ (one-tailed), $r = 0.19$.

[16]Whether the plaintiff's expert witness was a blind expert had a significant effect on his credibility as perceived by participants, $F(2, 272) = 28.64$, $p < 0.001$. Planned contrasts revealed a significant increase in overall credibility of the plaintiff's expert in the PLAINTIFF BE condition ($M = 6.94$, $SD = 1.28$) relative to the NEITHER BE condition ($M = 6.94$, $SD = 1.17$), $t(272) = 2.96$, $p < 0.001$ (one-tailed), $r = 0.18$. When only the defendant used a blind expert (DEFENDANT BE), the plaintiff's expert witness suffered a worse credibility assessment ($M = 5.73$, $SD = 1.14$) than in the NEITHER BE condition, $t(272) = 3.50$, $p < 0.001$ (one-tailed), $r = 0.21$.

for example, his technical training or intellect. The "genuineness" construct, in contrast, consisted of *honest, trustworthy,* and *fair,* and was meant to capture the believability of the witness.[17] *Skill* and *genuineness* scores are added together to give an overall *credibility* score. A factor analysis confirmed that credibility can be usefully conceptualized as composed of these two factors.[18]

Both constructs were significantly affected depending on whether a blind expert was used (skill: $F(2, 272) = 13.93$, $p < 0.001$; genuineness: $F(2, 272) = 37.77$, $p < 0.001$). However, as one might expect, the blind procedure had a greater impact on genuineness.[19] When an expert is blinded, the jury perceives greater skill and genuineness compared to the counterfactual situation where neither expert is blinded. But for the nonblinded expert in a blinded expert condition, the jury perceives the nonblinded expert as having lower genuineness but the same skill level as in the counterfactual condition.

E. Open-Ended Responses

After rendering their verdicts (but before being prompted with expert credibility questionnaires), participants were required to provide a sentence or two explaining their decisions. Participants were also allowed to provide any other comments at the very end of the study. All these responses were independently coded by two research assistants, with coding discrepancies resolved by the principal investigators.

---

[17]To appreciate the possible difference between these two factors, you might imagine, for example, an expert with outstanding credentials (but who seems to be lying). This expert would have a high skill score (he or she *could* give an accurate opinion), but a low genuineness score (since he or she lies). On the other hand, an expert might appear to be doing his or her honest best to give an accurate opinion, but nonetheless seem unqualified in training or intellect. This expert would have a high genuineness score, but a low skill score. The point is that both sorts of considerations (and possibly others) factor into an overall assessment of credibility, and different factors might be differentially sensitive to the blind expert concept.

[18]Principal component analysis converged after five iterations. Factor loadings were as follows:

	Plaintiff's Expert (Component)		*Defendant's Expert (Component)*	
	(1)	*(2)*	*(1)*	*(2)*
Trustworthy	0.956		0.907	
Honest	0.956		0.922	
Fair	0.833		0.934	
Clear		0.923		0.324
Logical		0.688		0.389
Knowledgeable		0.898		0.969

[19]Planned contrasts revealed that perceptions of the plaintiff's expert's skill, relative to the NEITHER BE condition ($M = 3.23$, $SD = 0.62$), significantly increased whenever he was the only blind expert ($M = 3.51$, $SD = 0.05$), $t(272) = 3.07$, $p = 0.002$, $r = 0.18$. When the defendant's expert was the only blind expert, however, perceptions of skill did not significantly change ($M = 3.10$, $SD = 0.53$), $t(272) = 1.36$, $p = 0.18$, $r = 0.08$. Assessments of genuineness also significantly increased in the PLAINTIFF BE condition ($M = 3.43$, $SD = 0.67$), relative to the NEITHER BE condition ($M = 3.16$, $SD = 0.65$), $t(272) = 2.48$, $p = 0.014$, $r = 0.15$. When the defendant's expert was the only blind expert, perceptions of the plaintiff's expert's genuineness significantly decreased ($M = 2.63$, $SD = 0.07$), $t(272) = 4.85$, $p < 0.001$, $r = 0.28$.

The participants' open-ended responses were largely driven by the facts in the case, as one would expect. For example, one juror said: "While it is understandable to think now that the Defendant should have done the MRI that would have identified the problem, it [is] also understandable to think that he would be conservative about his diagnosis and choose instead to continue observation instead of charging his patient for a test that may not be necessary." Others invoked policy considerations and mentioned particular facts, for example: "These kinds of lawsuits drive up medical costs for everyone, [and] the doctor met the standard of care based on the information he received from the patient and the tests in his office. Plus the patient did not go back sooner than 3 months." These themes about the standard of care were raised by the experts and attorneys throughout the trial.

Of those who were in blind expert conditions, 43 (20 percent) explicitly mentioned the blind procedure or the blind witness favorably (including those referring to the expert being picked from a "pool" or referring to the other expert as "hand-picked" or "cherry-picked"). Eleven more participants (5 percent) referred to the "bias" of the unblind expert, without referring to him as such, which totals 25 percent of the subjects who invoked blinding or witness bias as a basis for their decision. For example, one wrote that "the plaintiff's expert witness, being a blind witness, had no reason to be swayed." Another argued the facts of the case and then said, "plus, having a blind expert does make me feel more confident in Dr. Dennis' [the defendant's] decisions." Some seemed to think that this was a determinative factor, writing that "it's clear to me that the plaintiff's doctor has been purchased and will be biased to say whatever he needs in order to get his money. His opinion is worth little to nothing to me, whereas the integrity of the blind witness is inherently honest." Others mentioned the blind procedure only to say that it was "not the determining factor in my decision" or that "it did not sway me either way."

IV. Discussion and Limitations

The primary question for this study was whether an expert's testimony would be perceived as more credible, and thereby more strongly influence juror verdicts, if he or she was blinded. If so, then litigants would have an incentive to use the procedure, and it would actually have some chance of improving litigation, better and more legitimately serving its deterrence, compensation, and punishment functions.

The results indicate that using a blind expert provides a significant and meaningful benefit in terms of achieving a verdict favorable for one's client. The plaintiff in this case was at a disadvantage when neither expert was blinded—such as in a typical trial—winning only 45.9 percent of the verdicts. However, when the plaintiff used a blind expert, he was able to overcome the initial disadvantage and secure a majority of the verdicts (62.3 percent) from the jurors. A similar movement was observed in the other direction when the defendant's expert was blinded. Collectively, this amounted to a doubling of the odds of winning, which may be worth hundreds of thousands of dollars or more, depending on the damages at stake in the case (see Figure 1). It is worth emphasizing that the actors and

substantive testimony of the case were identical across conditions, and that participants were randomly assigned to experimental conditions, including a control. Thus, this finding allows a strong inference of causality.

In the real world, the impact of blinding may be even greater if litigants use blind opinions as screening devices for determining which cases to litigate and which to drop (or settle quickly). If blind experts are more objective and accurate, then the cases for which they provide favorable opinions will likely have other evidentiary features that cohere with the expert testimony, thereby strengthening the persuasiveness of the expert and otherwise enhancing the chance of winning. This "selection bias" toward meritorious cases was controlled for in the present experiment, since the facts and substantive testimony were the same across conditions, and blind experts were assigned to either side for experimental purposes. A litigant's strategy of obtaining early blind opinions and then litigating favorable cases may be even more lucrative than the $100,000 to $150,000 benefit calculated above (see Figure 1).

Thus, it may turn out that the blind expert creates a new equilibrium, wherein each party invests in purchasing a blinded expert opinion. In bona fide fields of scientific knowledge, one would expect that blinded experts will often agree with each other, since their opinions will not be polarized by the influence of litigants. If that is true, and if litigants are effectively limited to a single blinded opinion (Robertson 2010:212), then in most cases, a single blinded expert will actually appear for trial, since the other blinded expert's opinion will be hidden in attorney work-product protections (Robertson 2010:211). Both sides will bring blinded experts to trial only in the rare cases where the two blinded experts disagree, and the disagreement happens to be distributed in a way that is favorable to each sponsor (Robertson 2010:217). We would hypothesize that when jurors are presented with two blinded experts, they decide the case similarly as if there are two unblinded experts. This rare outcome would thus impose costs on the litigants, without providing a relative advantage. Future studies could explore this possibility. Rational litigants would need to account for such a potentially wasteful outcome in evaluating the net benefits of the procedure. (Robertson 2010:256–57, providing an equation for that purpose, and showing that on a wide range of assumptions, the tactic remains a rational strategy).

This experiment did not include a deliberation phase or allow jurors to vote in groups to render collective judgments, which would be the ideal dependent variable of interest. However, the six-level juror certainty data provide a useful proxy. One can imagine a sample of 12 jurors being drawn from the observed jurors in each group, and one might further hypothesize that jurors who have very strong personal verdicts will have a greater impact on the collective jury decisions. In the condition with no blind expert, with substantial portions across all six levels of the Likert scale, any given sample could be skewed either way; as such, both parties face substantial risk of unfavorable ultimate decisions. The blind expert conditions, in contrast, pushed jurors to higher levels of certainty in accordance with whichever side brought the blind expert. In these conditions, the party with the blind expert is likely to receive very favorable initial votes from a sample of 12 jurors, and also has some assurance that the outcome of deliberations will reflect this initial distribution. Thus, the predictability of jury decisions may be improved, which would be an important benefit for litigators (Jacobson et al. 2011).

We predicted that the effects on verdict would be mediated by perceptions of credibility. To explore this possibility, participants were asked to rate both expert witnesses on six factors from which a composite score of overall credibility was constructed. As hypothesized, credibility was significantly enhanced when the expert was described as a blind expert, and this was true for both the plaintiff's expert (6.39 to 6.94) and the defendant's expert (5.96 to 6.28).

Not only did blinding an expert boost the credibility of that expert, but it also *harmed* the credibility of the *opposing* expert. Apparently, participants perform a contrast between the two experts to determine their relative credibility. Thus, when the plaintiff retained the blind expert, the defendant's expert suffered a reduction in credibility from 5.96 to 5.35; likewise, when the defendant retained the blind expert, the plaintiff's expert suffered a reduction in credibility from 6.39 to 5.73. Closer examination of the skill and genuineness subdivisions reveals that this effect is driven by altering the relative genuineness of the unblind expert, which reduced from 3.16 to 2.63 for the plaintiff's expert and from 2.87 to 2.44 for the defendant's expert. Assessments of skill, on the other hand, although being lower, were not significantly different. Litigants may exploit this effect by using their own blind expert to paint the adversary's expert as an unreliable "hired gun," notwithstanding whatever impressive credentials or technical abilities he or she may possess.

This loss in credibility may have larger consequences for jury trials, where litigants speak through their witnesses and their attorneys. Trial attorneys are advised: "Your case will be presented largely through witnesses. If the witnesses are not credible, you have no chance of establishing your own credibility" (Easton 1998:8). If jurors perceive that one party is taking special precautions to ensure that the evidence is fair and unbiased, that party may also be able to present other evidence in a more favorable light, getting the benefit of the doubt from jurors. Likewise, closing arguments from such attorneys may be more persuasive. Blinding may fit into the larger narrative of a trial, showing which side is taking appropriate precautions and acting reasonably, versus which side is acting negligently or even recklessly in the light of known problems, such as witness bias (Ball & Keenan 2009).

These conjectures about spill-over effects are fodder for future research. However, the observed pain and suffering awards in this study are suggestive of such a spill-over effect. Using a blind expert may have an impact on the amount of money that a juror deems appropriate, an effect that becomes significant when combined with the differences in win rates.

Jury instructions were included in two conditions under the hypothesis that if the judge acknowledged the blinding process as a specific factor for considering the credibility of the experts, especially immediately before asking the jurors to decide the case, then the effects of the blind expert would be more substantial. This hypothesis was not supported. If anything, there was a counteractive effect, with the benefit of the blind expert on verdict outcome being slightly (but statistically insignificantly) reduced whenever instructions were included. Judicial instructions are notorious for being poorly understood by jurors, though some prior research has documented that they can have a significant effect on outcomes when understood (see generally Lieberman & Sales 1999). The given instruction was not particularly strong or emphatic in its support of blinding, concluding merely that "you *may*

find the blind expert more credible and you *may* decide to give the testimony of that expert additional weight. *However*, the ultimate decision about the credibility of witnesses is yours" (emphasis added). Even though it comes just moments before the juror decides the case, this particular wording may have been ambiguous to a juror trying to determine whether to give more weight to the blind expert. If a judge were to provide a more ringing endorsement of the concept, it may drive jurors further toward the blinded expert. This intervention would, however, also risk an appellate court reversing for the trial court invading the province of the jury.

The important finding is that blinding drives a significant change in outcomes regardless of jury instructions. Finding that judicial intervention is unnecessary adds further credence to the litigant-driven model of blind expertise, wherein self-interested actors would use the procedure to improve the values of their own cases.

This experiment suffered from several limitations. First, the stimulus was condensed and artificial. Although we consulted with practicing physicians and experts on trial practice to design the 35-minute core stimulus video, it was not a real trial, and indeed was much shorter than a real medical malpractice trial, which can proceed for days or even weeks. The impact of blinding may be smaller in the context of a full trial if it is swamped by other evidence and argument. Relatedly, we tested a particular sort of medical malpractice case as a stimulus. Blind expertise may have a different impact on other sorts of medical malpractice cases, or even other sorts of civil or criminal litigation.[20]

Second, it is likely that if the trial stimulus had provided a more robust discussion of the pain and suffering issue, including requests by the attorney for a specific dollar amount, jurors may have anchored on that reference point. We may have thereby been able to reduce the amount of variance in our data.

Third, the portions of the video stimuli that discuss the blinding procedure were also truncated. For example, a complete discussion of the blind procedure would have included some explanation about whether the litigant could draw multiple times until he or she received a favorable blind opinion, and then only disclose that one opinion to the factfinder, thereby exploiting a concealed selection bias. In conception, this is not a problem because there is a waiver of attorney work-product protections that would force such a disclosure (Robertson 2010), thereby functionally limiting litigants to a single blind expert opinion on a given question. We thus omitted discussion of this dynamic for fear of needlessly complicating the jury's analysis. However, some of the jurors' open-ended responses raised this possibility of iterative selection. For example, one wrote, "I suspect that the prosecution 'randomly' selected several blind experts until they found one with the same opinion as the plaintiff." Another asked: "My question is[,] how many blind witnesses can you legally hire? Did he hire three others who disagreed and cherry-picked the one who saw it his way?" Another asked about multiple selection and said, "the unclearness on this point made me wonder if the emphasis on cherry-picking of the other expert was really such a big deal."

---

[20]Even though the causal impact in this case was observed in both directions, and of about the same size, it is possible that there will be ceiling and floor effects in other cases where the baseline evidence is extremely lopsided for one party or the other and the presence of a blind expert may be insignificant. However, it is doubtful that such highly skewed cases would be tried to a jury rather than settled by rational parties. See Vidmar (1995:175).

This uncertainty apparently reduced the effectiveness of the blinding intervention for some respondents. Future experiments (and actual litigators) should clarify and settle this issue for jurors, and thereby may find an even larger effect for the blind procedure.

Fourth, the study was conducted with human subjects online. Prior research has shown that "the population of Mechanical Turk is at least as representative of the U.S. population as traditional subject pools" (Paolacci et al. 2010:411). Known experimental results have been replicated using the Mturk population (Berinsky et al. 2012). Still, it is likely that Mturkers may be more easily distracted from the trial compared to real jurors, and may even provide junk responses.[21] Such problems would increase noise in the data. In the real world, the effect of blinded experts may be even greater.

A fifth and related limitation is that our sample was somewhat younger, more female, and more educated than a median U.S. juror, though our randomization succeeded in distributing these differences across conditions. Our regression analyses do not suggest that these differences were driving our results, though future studies should explore the possibility that more educated people better appreciate the blinding concept.

A sixth limitation is that we did not provide the jury with an opportunity to deliberate. The literature on the effect of deliberation is complex, but some research has shown that deliberation can increase polarization among jurors, causing more extreme outcomes (Schkade et al. 2002). If that dynamic applied here, then the effect of blinding may be even greater. Still, further experiments are underway to replicate these findings and investigate the causal mechanisms in more realistic settings.

Seventh, we used a sample size of 275 subjects to populate five conditions in a factorial design, and used a traditional 0.05 threshold for statistical significance. We conducted (and reported) multiple statistical tests in our data set, which increased the risk of false positives.

Notwithstanding these limitations, as a randomized controlled trial with 275 jury-eligible adults from a national sample, using real doctors as experts in a videotaped stimulus with proper jury instructions from a judge, the present experiment has indicia of reliability (Bornstein 1999). The main effects were observed on multiple dependent variables, including verdict, certainty, expert credibility, and verdict size, which together tell a coherent causal story.

V. Implications and Conclusions

Importantly, the stimulus video was designed with a right answer to the question of whether the doctor met the standard of care of the medical profession based on national practice guidelines (Chou 2007). Even though we did not provide this information to jurors, it provides a point of reference for analyzing jury performance. In the control condition that

---

[21]We included a quiz to check whether they watched the video, but decided not to utilize it to exclude respondents because we had no baseline for how much real jurors pay attention and thus no way to set a threshold for satisfaction. We were concerned that screening out those who were not paying attention might unduly favor our hypotheses tests.

simulates real-world trials with counterpoised unblinded experts, most jurors (54 percent) failed to reach that conclusion, on the basis of testimony of two well-qualified experts who attempted to explain the standard of care. Jurors did worse than coin-flipping, an important and disappointing finding for the debate over jury competence in the status quo (see, e.g., Ivkovic & Hans 2003; Hans & Eisenberg 2011).

This finding is strikingly similar to that reported in the subset of real-world medical malpractice trials, where scholarly reviewers independently determine that the truth favors the plaintiff, and yet the defendant won nonetheless. For example, Studdert and Mello (2007) found that in such cases, defendants won 57 percent of the time. While disappointing as a matter of policy, this finding suggests that our experimental stimulus may have ecological validity as a typical medical malpractice case, where juries often get it wrong.

Policymakers and advocates use such findings to argue for limits on the role of jurors in the U.S. legal system, and advance proposals to instead relegate the fact-finding function to other institutions, such as judges, specialized health courts, or ex ante regulators.[22] Still, somebody has to decide these sorts of cases, and "in complicated fields like DNA, epidemiology, or chemistry, judges are also laypersons" (Vidmar & Diamond 2001:1169). In theory, specialist courts or expert regulators could have more competence, but would suffer from other problems such as selection bias or regulatory capture. Such reforms also face difficult questions of political economy.

One might argue that jurors are simply doing what they are told in the epistemic situation they are given. It may be that when jurors are given only two counterpoised experts—who are likely biased by their sponsor's selection, affiliation, and compensation—jurors are left near epistemic equipoise. The judge instructs them to return a defense verdict, given the preponderance of evidence standard, and they comply. In this light, jurors may be highly functioning components of a dysfunctional system. Rather than changing the decisionmaker, it may be prudent to instead provide factfinders with better epistemic signals. Blinded experts may fit that bill.

The foregoing data show that blind expertise more than doubles a litigant's odds of winning, an effect much larger than we hypothesized, and likely larger than litigants would otherwise assume. If litigants have so far assumed that the effect would be little or nothing, that may explain why the blind procedure has not already been utilized in litigation.

The economic findings displayed in Figure 1—that the presence of a blind expert may be worth hundreds of thousands of dollars to litigants—suggest that there is a very large upside to this strategy. Of course, the vast majority of cases are never tried, but the amounts of settlements are largely based on predictions about potential trial outcomes that will ensue if settlement negotiations fail (Priest & Klein 1984). In addition to this benefit, a blind expert may also allow litigants to efficiently screen and select cases, thereby reducing litigation costs by causing earlier settlements—dynamics not studied here. A blind expert

---

[22]See, e.g., Huber (1985) and Struve (2004), discussing such proposals. For example, in the Fair and Reliable Medical Justice Act, S. 1337, 109th Cong. § 3(d)(4)(B) (2005), senators proposed the creation of health courts presided over by judges "with health care expertise," but who only need to "meet applicable State standards for judges." Such judges would likely need to rely on expert witnesses, which just replicates the problem of bias.

may also carry extra weight with the trial and appellate judges, thus improving a litigant's odds of prevailing on *Daubert* motions, summary judgment motions, and posttrial review of verdicts. It certainly would not hurt.

Litigants will weigh these advantages against the cost of paying in advance for an opinion that may or may not turn out to be helpful. Traditional attorney work-product protections allow that a litigant can procure a blind expert opinion without risk of it harming his or her case. The attorney can discard the opinion if he or she finds it unhelpful, or supplement the opinion with a traditional expert witness over whom the attorney can exert greater control (a strategy not tested in the present experiment). Thus, on net, the blind procedure seems to be a rational strategy for litigants—they will pay the cost of such a blinded opinion in order to potentially double their odds of winning, and thus improve the expected value of their settlements.

Whether blinded experts will improve the process and outcomes of litigation is another question. Of course, litigant-induced biases are not the only problems with expert testimony. However, arguably, to the extent that the litigant-induced biases of unblinded experts degrade the reliability of their testimony, these biases undermine the ability of factfinders to do their jobs. The blind procedure preserves the best parts of the adversarial system, including robust cross-examination, while avoiding the problems of litigant-induced bias. If blinded experts are more reliable guides for the factfinder, and jurors disproportionately rely on their testimony, then blinded experts will improve the litigation process, making it more legitimate and improving the quality of its compensation, deterrence, and punishment signals.

References

Ake v. Oklahoma (1985) 470 U.S. 68.

American Medical Association Advocacy Resource Center (2011) *Caps on Damages.* Available at <http://www.ama-assn.org/resources/doc/arc/capsdamages.pdf>.

Ball, D., & D. Keenan (2009) *Reptile: The 2009 Manual of the Plaintiff's Revolution.* Portland, OR: TrialGuides. Available at < http://www.shop.reptilekeenanball.com/>.

Berinsky, A. J., G. A. Huber, & G. S. Lenz (2012) "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk," 20 *Political Analysis* 351.

Bornstein, B. H. (1999) "The Ecological Validity of Jury Simulations: Is the Jury Still Out?" 23(1) *Law & Human Behavior* 75.

Brekke, N., & E. Borgida (1988) "Expert Psychological Testimony in Rape Trials: A Social-Cognitive Analysis," 55 *J. Personality & Social Psychology* 372.

Brodsky, S. L., M. P. Griffin, & R. J. Cramer, (2010) "The Witness Credibility Scale: An Outcome Measure for Expert Witness Research," 28 *Behavioral Sciences & the Law* 892.

Brodsky, S. L., T. M. Neal. R. J. Cramer, & M. H. Ziemke, (2009) "Credibility in the Courtroom: How Likeable Should an Expert Witness Be?" 37 *J. of the American Academy of Psychiatry Law* 525.

Carrington, P. D. (1992) "One Law: The Role of Legal Education in the Opening of the Legal Profession Since 1776," 44 *Florida Law Rev.* 501.

Cecil, J. S., & T. E. Willging. (1993) *Court-Appointed Experts: Defining the Role of Experts Appointed Under Federal Rule of Evidence 706.* Washington, DC: Federal Judicial Center. Available at <http://www.uchastings.edu/faculty-administration/faculty/adjuncts/class-websites/willging/docs/class%209/CL9FJC-court-experts-1993.pdf>.

Chou, R. (2007) "Diagnosis and Treatment of Low Back Pain: A Joint Clinical Practice Guideline from the American College of Physicians and the American Pain Society," *Annals of Internal Medicine* 147.

Cooper, J., & I. M. Neuhaus (2000) "The 'Hired Gun' Effect: Assessing the Effect of Pay, Frequency of Testifying, and Credentials on the Perception of Expert Testimony," 24 *Law & Human Behavior* 149.

Danaher, J. (2011) "Blind Expertise and the Problem of Scientific Evidence," 15 *International J. of Evidence & Proof* 207.

Davis, J. H. (1996) "Group Decision Making and Quantitative Judgments: A Consensus Model," in E. Witte & J. Davis (eds.), *Understanding Group Behavior: Consensual Action by Small Groups.* Mahwah, NJ: Erlbaum.

Diamond, S., & J. Casper (1992) "Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury," 26 *Law & Society Rev.* 513.

Diamond, S., B. Murphy, M. R. Rose, & J. B. Meixner (2011) "Damage Anchors on Real Juries," 8 *J. of Empirical Legal Studies* 148. Available at <http://ssrn.com/abstract=1883861>.

Easton, S. D. (1998). *How to Win Jury Trials: Building Credibility with Judges and Jurors.* Philadelphia, PA: ALI-ABA. Available at <http://books.google.com/books?id=HhTPqlA0l6MC&printsec=frontcover#v=onepage&q&f=false>.

Gitlin, J. N., L.L. Cook, O. W. Linton, & E. Garrett-Mayer (2004) "Comparison of 'B' Readers' Interpretations of Chest Radiographs for Asbestos Related Changes," 11 *Acad. Radiology* 843.

Glyptis, P. T., & D. S. Givens (2011) "Developing a True Blind Review," September *ACR Bulletin* 24.

Gross, S. R. (1991) "Expert Evidence," 1991 *Wisconsin Law Rev.* 1114.

Hans, V. P., & T. Eisenberg (2011) "The Predictability of Juries," *Cornell Legal Studies Research Paper.* Available at <http://ssrn.com/abstract=1686369>.

Huber, P. (1985) "Safety and the Second Best: The Hazards of, Public Risk Management in the Courts," 85 *Columbia Law Rev.* 277.

Hyams, A. L., J. A. Brandenburg, S. R. Lipsitz, D. W. Shapiro, & T. A. Brennan (1995) "Practice Guidelines and Malpractice Litigation: A Two-Way Street," 122(6) *Annals of Internal Medicine* 450.

Ivkovic, S. K., & V. P. Hans (2003) "Jurors' Evaluations of Expert Testimony: Judging the Messenger and the Message," 28(2) *Law & Social Inquiry* 441.

Jacobson, J., J. Dobbs-Marsh, V. Liberman, & J. A. Minson (2011) "Predicting Civil Jury Verdicts: How Attorneys Use (and Misuse) a Second Opinion," 8 *J. of Empirical Legal Studies* 99.

Johnson, M. T., C. Krafka, & J. S. Cecil (2000) *Expert Testimony in Federal Civil Trials: A Preliminary Analysis.* Washington, DC: Federal Judicial Center. available at <http://www.fjc.gov/public/pdf.nsf/lookup/ExpTesti.pdf/$file/ExpTesti.pdf>.

Kahneman, D., D. Schkade, & C. R. Sunstein (1998) "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages," 16 *J. Risk & Uncertainty* 49.

King, G., & M. Roberts *How Robust Standard Errors Expose Methodological Problems They Do Not Fix.* Available at <http://gking.harvard.edu/files/robust.pdf>.

Langbein, J. H. (1985) "The German Advantage in Civil Procedure," 52 *Univ. of Chicago Law Rev.* 823.

Lieberman, J. D., & B. D. Sales (1999) "The Effectiveness of Jury Instructions," in W. F. Batt & J. Abbott (eds.) *Handbook of Jury Research,* §18. Philadelphia, PA: American Law Institute.

Liptak, A. (2008) "Experts Hired to Shed Light Can Leave U.S. Courts in Dark," August 12 *New York Times* A1.

Mid-State Fertilizer Co. v. Exchange Nat'l Bank of Chi. (7th Cir. 1989) 877 F.2d 1333 (Easterbrook, J.).

Oliver, L. C., L. S. Welch, & M. R. Harbut (2004) "Letter to the Editor," 11 *Acad. Radiology* 1397.

Otto, R. K. (1989) "Bias and Expert Testimony of Mental Health Professionals in Adversarial Proceedings: A Preliminary Investigation," 7 *Behaviorial Science & Law* 267.

Pace, N. M., D. Golinelli, & L. Zakaras (2004) *Capping Non-Economic Awards in Medical Malpractice Trials.* Available at <http://www.rand.org/pubs/monographs/2004/RAND_MG234.pdf>.

Paolacci, G., J. Chandler, & P. G. Ipeirotis (2010) "Running Experiments on Amazon Mechanical Turk," 5(5) *Judgment & Decision Making* 411. Available at <http://ssrn.com/abstract=1626226>.

Priest, G. L., & B. Klein (1984) "The Selection of Disputes for Litigation," 13 *J. of Legal Studies* 4.

Risinger, M. D. (2009) "The NAS/NRC Report on Forensic Science: A Glass Nine-Tenths Full (This Is About the Other Tenth)," 50 *Jurimetrics J.* 21.

Risinger, M. D., M. J. Saks, W.C. Thompson, & R. Rosenthal (2002) "The Daubert/Kumho Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion," 90 *California Law Rev.* 1.

Robertson, C. T. (2010) "Blind Expertise," 85 *New York Univ. Law Rev.* 175.

Rosenthal, R. (1978) "How Often Are Our Numbers Wrong?" 33 *American Psychologist* 1005.

Saks, M. J., L. Hollinger, R. Wissler, D. Evans, & A. J. R. Hart (1997) "Reducing Variability in Civil Jury Awards," 21 *Law & Human Behavior* 243.

Sanders, J. (1993) "Jury Deliberation in a Complex Case: *Havner v. Merrell Dow Pharmaceuticals,*" 16 *Justice System J.* 45.

Schkade, D. A., D. Kahneman, & C. Sunstein, "Deliberating About Dollars: The Severity Shift," in C. R. Sunstein, R. Hastie, J. W. Payne, D. A. Schkade, & W. K. Viscusi (eds.), *Punitive Damages: How Juries Decide,* pp.43–61.

Selvin, M., & L. Picus (1987) *The Debate Over Jury Performance: Observations From A Recent Asbestos Case.* RAND Corp. Available at <http://www.rand.org/pubs/reports/2007/R3479.pdf>.

Semelka, R. C., A. F. Ryan, S. Yonkers, & L. Braga (2010) "Objective Determination of Standard of Care: Use of Blind Readings by External Radiologists," 195 *American J. of Radiology* 429.

Sheppard, B. H. & N. Vidmar (1980) "Adversary Pretrial Procedures and Testimonial Evidence: Effects of Lawyer's Role and Machiavellianism," 39 *J. of Personality & Social Psychology* 320.

Shuman, D. W., E. Whitaker, & A. Champagne (1994) "An Empirical Examination of the Use of Expert Witnesses in the Courts—Part 2: A Three City Study," 3 *Jurimetrics J.* 193.

Snodgrass, R. (2006) "Single- Versus Double-Blind Reviewing: An Analysis of the Literature," 35 *Sigmod Record* 8.

Snyder, B. (2000) "Protecting the Media from Excessive Damages: The Nineteenth-Century Origins of Remittitur and its Modern Application in *Food Lion,*" 24 *Vermont Law Rev.* 299.

Solum, L. B. (2004) "Procedural Justice," 78 *Southern California Law Rev.* 181.

Struve, C. T. (2004) "Doctors, the Adversary System, and Procedural Reform in Medical Liability Litigation," 72 *Fordham Law Rev.* 943.

Studdert, D. M., & M. M. Mello (2007) "When Tort Resolutions are 'Wrong': Predictors of Discordant Outcomes in Medical Malpractice Litigation," 36 *J. of Legal Studies* S47.

Vidmar, N. (1995) *Medical Malpractice and the American Jury.* Ann Arbor, MI: Univ. of Michigan Press.

Vidmar, N., & S. S. Diamond (2001) "Juries and Expert Evidence," 66 *Brooklyn Law Rev.* 1121.